时空模型的局部众数回归

汪红霞, 林金官*, 黄性芳

南京审计大学统计与数学学院统计学系, 南京 211815

E-mail: hxwang@nau.edu.cn, jglin@nau.edu.cn, xfhuang@nau.edu.cn

收稿日期: 2018-07-30; 接受日期: 2019-01-06; 网络出版日期: 2019-11-08; * 通信作者
国家自然科学基金 (批准号: 17CTJ016) 资助项目

摘要 时空数据经常含有奇异点或来自重尾分布, 此时基于最小二乘的估计方法效果欠佳, 需要更稳健的估计方法. 本文提出时空模型的基于局部众数 (local modal, LM) 的局部线性估计方法. 理论和数据分析结果都显示, 若数据含有奇异点或来自重尾分布, 基于局部众数的局部线性方法比基于最小二乘的局部线性方法有效; 若数据无奇异点且来自正态分布, 两种方法效率渐近一致. 本文采用众数期望最大化 (modal expectation-maximization, MEM) 算法, 并在数据相依情形下得出估计量的渐近正态性.

关键词 时空模型 众数期望最大化 混合相依 局部线性回归

MSC (2020) 主题分类 62G05, 62G08

1 引言

时空数据统计分析在环境科学、地球科学、医药科学和经济科学等许多领域中有重要应用. 时空数据两大显著特点是相关性和异质性. 异质性一般用“局部”的方法来刻画, 例如, 文献 [1, 2] 采用局部线性方法来刻画空间数据的异质性; 文献 [3, 4] 采用局部线性方法来刻画时空数据的异质性. 这些局部方法都是基于最小二乘的估计方法, 最小二乘估计方法有很多很好的性质, 尤其当数据来自正态分布. 但若数据含有奇异点或来自重尾分布, 则基于最小二乘的方法不够稳健. 文献 [5] 提出了更为稳健的空间数据分位数回归; 文献 [6] 提出了时空数据的基于最小一乘的局部线性方法. 然而, 这些方法在数据没有奇异点且来自正态总体时会损失一些效率.

文献 [7] 提出了基于局部众数回归的新的估计方法. 局部众数方法通过增加一个调节参数, 自适应选择观测数据, 同时达到稳健性和有效性. 局部众数估计方法不仅在数据有奇异点或来自重尾分布时有较好的稳健性, 而且在数据无奇异点且误差分布是正态分布时, 效率与最小二乘估计方法是渐近一致的. 众数回归引起了学者们的广泛关注, 文献 [8–10] 分别推广众数回归到单指标模型、部分线性变系数模型和部分线性单指标模型. 文献 [11] 提出了基于局部众数的非线性模型稳健的变量选择方

英文引用格式: Wang H X, Lin J G, Huang X F. Local modal regression for the spatio-temporal model (in Chinese). *Sci Sin Math*, 2021, 51: 615–630, doi: 10.1360/N012018-00191

法. 文献 [12, 13] 分别利用局部众数回归分析饮食数据和预测阿尔茨海默病. 文献 [14] 提出了基于核密度估计的非参数众数回归. 文献 [15] 回顾了利用核密度估计的众数回归的最新进展.

基于局部众数方法有很好的性质, 本文将局部众数方法推广到如下非参数时空模型:

$$Y_{i,t} = m(X_{i,t}) + R_{i,t}, \quad (1.1)$$

其中 $Y_{i,t} \in \mathbb{R}$, $X_{i,t} \in \mathbb{R}^d$, $m(\cdot)$ 是未知的趋势函数, $R_{i,t}$ 是随机误差, 满足 $E(R_{i,t} | X_{i,t} = \mathbf{x}) = 0$, $\text{Var}(R_{i,t} | X_{i,t} = \mathbf{x}) = \sigma^2(\mathbf{x})$, $\mathbf{i} \in \Lambda_{\mathbf{n}} = \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_N\}$, $t \in T_{\mathbf{n}} = \{1, \dots, n_{N+1}\}$, $\mathbf{n} = (n_1, \dots, n_{N+1})$, $n^* = n_1 \times \dots \times n_{N+1}$. 非参数时空模型已有学者研究 (参见文献 [16]). 目前, 局部众数方法未应用到非参数时空模型, 本文希望在以下三方面作一些改进. (1) 推广局部众数方法到非参数时空模型 (1.1), 期望同时达到稳健性和有效性. 理论、模拟和实例结果显示, 若数据含有奇异点或来自重尾分布, 基于局部众数的局部线性方法比基于最小二乘的局部线性方法有效; 若数据无奇异点且来自正态分布, 两种方法效率渐近一致. (2) 模型更贴近实际数据, 同时考虑异质性和相依性, 本文考虑非参数模型且假设误差服从 α 混合相依. (3) 在较弱的条件下得到估计量的大样本性质, 为后续统计推断提供理论依据. 本文没有假设 $R_{i,t}$ 关于时间和空间是严平稳的, 只是假设在固定空间点 $\mathbf{i}_0$, $\{R_{\mathbf{i}_0,t}, t \in T_{\mathbf{n}}\}$ 是严平稳的, 在固定时间点 t_0 , $\{R_{i,t_0}, \mathbf{i} \in \Lambda_{\mathbf{n}}\}$ 是严平稳的.

本文结构如下: 第 2 节给出基于局部众数的最小线性估计方法、众数期望最大化算法、估计量的理论性质和最优参数的选取. 第 3 节给出模拟和实例. 第 4 节给出理论证明.

2 局部众数回归估计量

由 Taylor 展式, 若 $\mathbf{x}$ 是 $\mathbf{x}_0$ 邻域的点, 则近似有

$$m(\mathbf{x}) \approx m(\mathbf{x}_0) + [m'(\mathbf{x}_0)]^\top (\mathbf{x} - \mathbf{x}_0) \equiv \beta_0 + \beta_1^\top (\mathbf{x} - \mathbf{x}_0),$$

其中 $\beta_0 = m(\mathbf{x}_0)$, $\beta_1 = m'(\mathbf{x}_0)$. 为方便起见, 记 $\boldsymbol{\theta} = (\beta_0, \beta_1^\top)^\top$, 利用局部众数方法来估计 $\boldsymbol{\theta}$, 即极大化下面的目标函数:

$$l(\boldsymbol{\theta}) = \frac{1}{n^*} \sum_{\mathbf{i} \in \Lambda_{\mathbf{n}}} \sum_{t \in T_{\mathbf{n}}} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi_{h_2}(Y_{i,t} - \beta_0 - \beta_1^\top (X_{i,t} - \mathbf{x}_0)), \quad (2.1)$$

其中 $\mathcal{K}_{h_1}(\mathbf{t}) = \frac{1}{h_1^d} \mathcal{K}(\frac{\mathbf{t}}{h_1})$, $\phi_{h_2}(t) = \frac{1}{h_2} \phi(\frac{t}{h_2})$, $\mathcal{K}(\cdot)$ 是 d 维核概率密度函数, $\phi(\cdot)$ 是一维核密度函数. h_1 和 h_2 是带宽. 核函数的选取不是很重要, 故为了计算方便, 本文中选取 $\phi(\cdot)$ 为 Gauss 核. h_1 和 h_2 的选取很重要, 我们将在后面的章节介绍. 记 $l(\boldsymbol{\theta})$ 的极大值点为 $\hat{\boldsymbol{\theta}}$, 则

$$\hat{m}(\mathbf{x}_0) = \hat{\beta}_0, \quad \hat{m}'(\mathbf{x}_0) = \hat{\beta}_1.$$

由于 (2.1) 的极大值点没有显式表达式, 本文采用 MEM 算法来计算 $\hat{\boldsymbol{\theta}}$. 为了书写方便, 我们把 $\{(X_{i,t}, Y_{i,t}), \mathbf{i} \in \Lambda_{\mathbf{n}}, t \in T_{\mathbf{n}}\}$ 按任意顺序排列, 记为 $\{(X_j, Y_j), j = 1, 2, \dots, n^*\}$.

2.1 MEM 算法

本节将推广文献 [17] 介绍的 MEM 算法来计算 (2.1) 的极大值点. 求 (2.1) 的极大值点, 即求下式的极大值点:

$$\ln \sum_{j=1}^{n^*} \mathcal{K}_{h_1}(X_j - \mathbf{x}_0) \phi_{h_2}(Y_j - \beta_0 - \beta_1^\top (X_j - \mathbf{x}_0)).$$

为了构造期望, 引入 $\pi(j | \boldsymbol{\theta}^{(k)})$, 上式等价于下式. 由 Jensen 不等式, 有

$$\begin{aligned} & \ln \sum_{j=1}^{n^*} \pi(j | \boldsymbol{\theta}^{(k)}) \frac{\mathcal{K}_{h_1}(X_j - \mathbf{x}_0) \phi_{h_2}(Y_j - \beta_0 - \beta_1^\top (X_j - \mathbf{x}_0))}{\pi(j | \boldsymbol{\theta}^{(k)})} \\ & \geq \sum_{j=1}^{n^*} \pi(j | \boldsymbol{\theta}^{(k)}) \ln \frac{\mathcal{K}_{h_1}(X_j - \mathbf{x}_0) \phi_{h_2}(Y_j - \beta_0 - \beta_1^\top (X_j - \mathbf{x}_0))}{\pi(j | \boldsymbol{\theta}^{(k)})}. \end{aligned}$$

考虑 Jensen 不等式成立条件, 与一般的期望最大化 (expectation-maximization, EM) 算法类似, MEM 算法也包含两步: E 步和 M 步. 令 $\boldsymbol{\theta}^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)\top})^\top$ 为初始点, 由 $k = 0$ 开始.

E 步: 更新 $\pi(j | \boldsymbol{\theta}^{(k)})$ 为

$$\pi(j | \boldsymbol{\theta}^{(k)}) = \frac{\mathcal{K}_{h_1}(X_j - \mathbf{x}_0) \phi_{h_2}(Y_j - \beta_0^{(k)} - \beta_1^{(k)\top} (X_j - \mathbf{x}_0))}{\sum_{i=1}^{n^*} \mathcal{K}_{h_1}(X_i - \mathbf{x}_0) \phi_{h_2}(Y_i - \beta_0^{(k)} - \beta_1^{(k)\top} (X_i - \mathbf{x}_0))}, \quad j = 1, 2, \dots, n^*.$$

M 步: 更新 $\boldsymbol{\theta}^{(k+1)}$ 为

$$\boldsymbol{\theta}^{(k+1)} = \arg \max_{\boldsymbol{\theta}} \sum_{j=1}^{n^*} [\pi(j | \boldsymbol{\theta}^{(k)}) \ln \phi_{h_2}(Y_j - \beta_0 - \beta_1^\top (X_j - \mathbf{x}_0))] = (\mathcal{X}^\top W_k \mathcal{X})^{-1} \mathcal{X}^\top W_k \mathcal{Y}, \quad (2.2)$$

其中

$$\mathcal{X} = \begin{pmatrix} 1 & (X_1 - \mathbf{x}_0) \\ \vdots & \vdots \\ 1 & (X_n - \mathbf{x}_0) \end{pmatrix}, \quad W_k = \begin{pmatrix} \pi(1 | \boldsymbol{\theta}^{(k)}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \pi(n^* | \boldsymbol{\theta}^{(k)}) \end{pmatrix}, \quad \mathcal{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{n^*} \end{pmatrix}.$$

迭代直到 $\boldsymbol{\theta}^{(k)}$ 收敛为止.

MEM 算法的性质可参见文献 [17], 若 $h_2 \rightarrow \infty$, 易见 E 步为

$$\pi(j | \boldsymbol{\theta}^{(k)}) \rightarrow \frac{\mathcal{K}_{h_1}(X_j - \mathbf{x}_0)}{\sum_{i=1}^{n^*} \mathcal{K}_{h_1}(X_i - \mathbf{x}_0)} \propto \mathcal{K}_{h_1}(X_j - \mathbf{x}_0),$$

则由 (2.2) 知, 局部众数回归收敛于普通的局部线性回归, 即局部线性回归是局部众数回归的极限形式. 由 MEM 算法可知, 时空数据的局部线性回归和局部众数回归的最大的不同点在 E 步. 局部线性回归中 $(\mathbf{X}_i, Y_i)$ 的贡献取决于权 $\mathcal{K}_{h_1}(X_i - \mathbf{x}_0)$, 只考虑了 X_i 和 $\mathbf{x}_0$ 的接近程度. 而局部众数回归中, 不仅考虑了 X_i 和 $\mathbf{x}_0$ 的接近程度, 也考虑了 Y_i 和回归曲线的接近程度, 对那些离回归曲线较远的点赋予较小的权, 从而达到自适应的目的.

2.2 理论性质

为了得到渐近性质, 假设对于固定时刻 t_0 , $\{R_{i,t_0}, i \in \Lambda_n\}$ 满足如下混合条件: 存在一函数 $\varphi_1(u) \searrow 0$, 当 $u \rightarrow \infty$, $\varphi_1(0) = 1$, 并且 $E, E' \subset \Lambda_n$ 时, 有

$$\begin{aligned} \alpha(\beta(E), \beta(E')) &= \sup\{P(AB) - P(A)P(B) \mid A \in \beta(E), B \in \beta(E')\} \\ &\leq \psi_1(\text{Card}(E), \text{Card}(E')) \varphi_1(d(E, E')), \end{aligned} \quad (2.3)$$

其中 $\beta(E)$ 表示由 $\{R_{i,t'}\}_{i \in E}$ 产生的 Borel σ -域, $\text{Card } E$ 表示集合 E 中元素的个数, $d(E, E')$ 为 Euclid 距离, $\psi_1: \mathbb{N}^2 \rightarrow \mathbb{R}^+$ 为对称非减正函数.

类似地, 假设对任一固定位置 $\mathbf{i}_0$, $\{R_{\mathbf{i}_0,t}, t \in T_n\}$ 满足如下混合条件: 存在函数 $\varphi_2(u) \searrow 0$, 当 $t \rightarrow \infty$, $\varphi_2(0) = 1$, 并且 $G, G' \subset T_n$ 时, 有

$$\begin{aligned} \alpha(\beta(G), \beta(G')) &= \sup\{P(AB) - P(A)P(B) \mid A \in \beta(G), B \in \beta(G')\} \\ &\leq \psi_2(\text{Card}(G), \text{Card}(G'))\varphi_2(d(G', G)), \end{aligned} \quad (2.4)$$

其中 $\beta(G)$ 表示由 $\{R_{\mathbf{i}_0,t}\}_{t \in G}$ 产生的 Borel σ -域, $\psi_2: \mathbb{N}^2 \rightarrow \mathbb{R}^+$ 为一对称非减函数.

下面的假设对于推导估计量的渐近性质是必要的, 从现在起, c 表示常数, 在不同行值可能不一样.

假设 2.1 (A1) 随机域 $\{X_{\mathbf{i},t}\}$ 是严平稳的. 对于 $\Lambda_n \times T_n$ 中不同的点 $(\mathbf{i}, t)$ 和 $(\mathbf{j}, \tau)$, 向量 $X_{\mathbf{i},t}$ 和 $X_{\mathbf{j},\tau}$ 有联合概率密度函数 $f_{(\mathbf{i},t)(\mathbf{j},\tau)}$. f 记为 $X_{\mathbf{i},t}$ 的边缘概率密度函数, 在 $\mathbf{x}_0$ 点处有连续的一阶导数且 $f(\mathbf{x}_0) > 0$.

(A2) $m(\mathbf{x})$ 二阶可导, 记 $m'(\mathbf{x})$ 和 $m''(\mathbf{x})$ 为其在点 $\mathbf{x}$ 处的梯度和二阶导矩阵且 $m''(\mathbf{x})$ 在所有 $\mathbf{x} \in \mathbb{R}^d$ 处连续.

(A3) $F(\mathbf{x}, h_2)$ 、 $G(\mathbf{x}, h_2)$ 、 $H(\mathbf{x}, h_2)$ 、 $E[\phi'_{h_2}(R_{\mathbf{i},t})^3 \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}]$ 和 $E[\phi'''_{h_2}(R_{\mathbf{i},t}) \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}]$ 在 $\mathbf{x}_0$ 点处关于 $\mathbf{x}$ 连续, 其中 $F(\mathbf{x}, h_2) = E[\phi''_{h_2}(R_{\mathbf{i},t}) \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}]$, $G(\mathbf{x}, h_2) = E[\phi'^2_{h_2}(R_{\mathbf{i},t}) \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}]$, $H(\mathbf{x}, h_2) = E[\phi''^2_{h_2}(R_{\mathbf{i},t}) \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}]$.

(A4) $\mathcal{K}(\cdot)$ 是对称的概率密度函数, 支集为 $[-1, 1]^d$.

(A5) $E[\phi'_{h_2}(R_{\mathbf{i},t}) \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}] = 0$.

(A6) $n^* \rightarrow \infty$ 时, $h_1 \rightarrow 0$, $(n^*h_1^d)^3 h_1^{4d} \rightarrow \infty$, 而 h_2 是不依赖于 n^* 的常数.

(A7) $\sum_{i=1}^{\infty} i^{N-1} \varphi_1(i) < \infty$, $\sum_{i=1}^{\infty} \varphi_2(i) < \infty$.

(A8) 对于任意 $h_2 > 0$, 有 $F(\mathbf{x}, h_2) < 0$.

注 2.1 (A1) 和 (A7) 与文献 [18] 中假设类似. (A2)–(A6) 和 (A8) 与文献 [7] 中假设类似. (A4) 只是为了证明方便, 可以放松到支集为 $[-M, M]^d$, M 为某个常数. 公式 (1.1) 中对于误差项的假设 $E(R_{\mathbf{i},t} \mid \mathbf{X}_{\mathbf{i},t} = \mathbf{x}) = 0$ 和 (A5) 在给定 $\mathbf{X}_{\mathbf{i},t} = \mathbf{x}$ 条件下, 当 $R_{\mathbf{i},t}$ 的条件概率密度函数对称时是可以同时满足的. 从后面 h_2 的选取方法会发现, 最优的 h_2 与 n^* 无关, 所以, (A6) 假设 h_2 是不依赖于 n^* 的常数.

下面给出主要结果, 证明将在第 4 节给出.

定理 2.1 假设 (A1)–(A8) 成立, 估计量有如下收敛速度:

$$|\hat{m}(\mathbf{x}_0) - m(\mathbf{x}_0)| = O_p((nh_1^d)^{-\frac{1}{2}} + h_1^2), \quad |h_1(\hat{m}'(\mathbf{x}_0) - m'(\mathbf{x}_0))| = O_p((nh_1^d)^{-\frac{1}{2}} + h_1^2) \cdot \mathbf{1}.$$

为了书写方便, 记 $\boldsymbol{\theta}^* = (\beta_0, h_1 \boldsymbol{\beta}_1^\top)^\top$, $\hat{\boldsymbol{\theta}}^* = (\hat{\beta}_0, h_1 \hat{\boldsymbol{\beta}}_1^\top)^\top$,

$$\begin{aligned} \mathbf{c}_1 &= \begin{pmatrix} \frac{1}{2} \int_{\mathbb{R}^d} \mathcal{K}(\mathbf{u}) \mathbf{u}^\top m''(\mathbf{x}_0) \mathbf{u} d\mathbf{u} \\ \frac{1}{2} \int_{\mathbb{R}^d} \mathbf{u} \mathcal{K}(\mathbf{u}) \mathbf{u}^\top m''(\mathbf{x}_0) \mathbf{u} d\mathbf{u} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) d\mathbf{u} & \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u}^\top d\mathbf{u} \\ \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u} d\mathbf{u} & \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u} \mathbf{u}^\top d\mathbf{u} \end{pmatrix}, \\ \mathbf{U} &= \begin{pmatrix} f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathcal{K}(\mathbf{u}) d\mathbf{u} & f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathbf{u}^\top \mathcal{K}(\mathbf{u}) d\mathbf{u} \\ f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathbf{u} \mathcal{K}(\mathbf{u}) d\mathbf{u} & f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathbf{u} \mathbf{u}^\top \mathcal{K}(\mathbf{u}) d\mathbf{u} \end{pmatrix}. \end{aligned}$$

定理 2.2 假设 (A1)–(A7) 成立, 则估计量的偏差和方差如下:

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) &= h_1^2 f(\mathbf{x}_0) \mathbf{U}^{-1} \mathbf{c}_1 (1 + o(1)), \\ \text{Var}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) &= (n^* h_1^d)^{-1} F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top (1 + o(1)). \end{aligned}$$

定理 2.3 假设 (A1)–(A7) 成立, 估计量有如下渐近正态性:

$$(n^* h_1^d)^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^* - \mathbf{E}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)) \xrightarrow{\mathcal{L}} N(0, F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top).$$

类似文献 [1, 定理 3.1], 结合本文定理 2.3, 易知非参数时空模型基于最小二乘的局部线性回归估计量的偏差和基于局部众数回归估计量的偏差一样, 但局部线性回归估计量的方差为

$$(n^* h_1^d)^{-1} \sigma^2(\mathbf{x}_0) f(\mathbf{x}_0) \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top.$$

为了比较两种估计方法的效率, 定义相对效率 $R_e(\mathbf{x}_0, h_2)$ 为两者方差之比

$$R_e(\mathbf{x}_0, h_2) = \frac{\text{Var}(LM)}{\text{Var}(L_2)} = \frac{G(\mathbf{x}_0, h_2) F(\mathbf{x}_0, h_2)^{-2}}{\sigma^2(\mathbf{x}_0)}.$$

关于相对效率有如下定理.

定理 2.4 令 $g_{R|X_{i,t}}(t | \mathbf{x})$ 为给定 $X_{i,t} = \mathbf{x}$ 时 $R_{i,t}$ 的条件概率密度函数. 对于任意给定的 $\mathbf{x}$, 有以下结果:

- (1) $\lim_{h_2 \rightarrow \infty} R_e(\mathbf{x}_0, h_2) = 1$, 故 $\inf_{h_2} R_e(\mathbf{x}_0, h_2) \leq 1$.
- (2) 若 $g_{R|X_{i,t}}(t | \mathbf{x})$ 是正态概率密度函数, 对于任意 h_2 , 有 $R_e(\mathbf{x}_0, h_2) > 1$, 但 $\inf_{h_2} R_e(\mathbf{x}_0, h_2) = 1$.
- (3) 假设 $g_{R|X_{i,t}}(t | \mathbf{x})$ 的三阶导数有界, 若 $h_2 \rightarrow 0$, 则 $R_e(\mathbf{x}_0, h_2) \rightarrow \infty$.

注 2.2 定理 2.4 的证明与文献 [7, 定理 2.4] 的证明类似, 故本文忽略定理 2.4 的证明. 定理 2.4(1) 表明, 选择合适的 h_2 , 可以使局部众数回归的效率高于最小二乘回归的效率. 定理 2.4(2) 表明, 即使在 $X_{i,t} = \mathbf{x}$ 条件下, 误差的分布是正态分布, 局部众数回归的效率低于最小二乘回归的效率, 但可以选择合适的 h_2 , 使得局部众数回归的效率接近最小二乘回归的效率. 定理 2.4(3) 显示, h_2 不能太小, 否则局部众数回归的效率会很差, 故合理选择参数很重要. 接下来介绍参数的选取.

2.3 最优参数选取

参数 h_1 和 h_2 的选取思路如下: 先基于最小二乘局部线性回归, 得到 h_1^{LLR} 和 $\hat{m}(X_{i,t})$, 进而得到 $\hat{R}_{i,t} = Y_{i,t} - \hat{m}(X_{i,t})$. 极小化 $\hat{R}_e(h_2)$ (具体表达式见下面步骤 (2)) 得到 h_2^{opt} . 比较基于最小二乘和局部众数的理论最优带宽, 得到 h_1^{opt} . 具体分以下三步:

(1) 基于最小二乘局部线性回归的最优带宽 h_1^{LLR} 的选取. 常用方法为文献 [19] 提出的交互验证 (cross validation, CV) 法, 极小化下面的目标函数:

$$h_1^{\text{LLR}} = \arg \min_{h_1} \sum_{i \in \Lambda_n} \sum_{t \in T_n} [m(X_{i,t}) - \hat{m}_{\neq(i,t)}(h_1)]^2,$$

其中 $\hat{m}_{\neq(i,t)}(h_1)$ 是剔除了点 (i, t) 的信息, 用剩下的 $n^* - 1$ 个点估计 $m(X_{i,t})$ 的估计值.

(2) h_2 的选取. $h_2^{\text{opt}} = \arg \min_{h_2} \hat{R}_e(h_2) = \arg \min_{h_2} \frac{\hat{G}(h_2) \hat{F}(h_2)^{-2}}{\hat{\sigma}^2}$, 其中 $\hat{\sigma}^2$ 是样本的方差,

$$\hat{F}(h_2) = \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \phi''_{h_2}(\hat{R}_{i,t}), \quad \hat{G}(h_2) = \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \{\phi'_{h_2}(\hat{R}_{i,t})\}^2.$$

根据文献 [7], 现实中一般用网格搜索法, $h_2 = 0.5\hat{\sigma} \times 1.02^j$, $j = 0, 1, \dots, 100$.

(3) 基于局部众数回归的最优带宽 h_1^{opt} 的选取. 由定理 2.2 可知, 局部众数回归估计量的均方误差为

$$(h_1^2 f(\mathbf{x}_0) e_1^\top \mathbf{U}^{-1} \mathbf{c}_1)^2 + (n^* h_1^d)^{-1} F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) e_1^\top \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top e_1,$$

其中 $e_1 = (1, 0, \dots, 0)^\top$ 是 $d+1$ 维列向量, 第一个分量为 1, 其余全为 0. 极小化均方误差, 易得 h_1 的理论最优为

$$h_1^{\text{opt}} = \left[\frac{d \cdot F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) e_1^\top \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top e_1}{4n^* f(\mathbf{x}_0) (e_1^\top \mathbf{U}^{-1} \mathbf{c}_1)^2} \right]^{\frac{1}{4+d}}.$$

而基于最小二乘局部线性回归的均方误差为关于

$$(h_1^2 f(\mathbf{x}_0) e_1^\top \mathbf{U}^{-1} \mathbf{c}_1)^2 + (n^* h_1^d)^{-1} \sigma^2(\mathbf{x}_0) f(\mathbf{x}_0) e_1^\top \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top e_1$$

求导, 得 h_1^{LLR} 的理论最优为

$$h_1^{\text{LLR}} = \left[\frac{d \cdot \sigma^2(\mathbf{x}_0) e_1^\top \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top e_1}{4n^* f(\mathbf{x}_0) (e_1^\top \mathbf{U}^{-1} \mathbf{c}_1)^2} \right]^{\frac{1}{4+d}}.$$

故实际上 $h_1^{\text{opt}} = \hat{R}_e(h_2^{\text{opt}})^{\frac{1}{4+d}} h_1^{\text{LLR}}$.

3 模拟和实例

3.1 模拟

本节将通过 4 个模拟展示我们方法的估计效果. 为简单起见, 令 $d = 1$, $N = 2$, 故考虑如下模型:

$$Y_{i,j,t} = 2 \sin(2\pi X_{i,j,t}) + R_{i,j,t}, \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2, \quad t = 1, \dots, n_3,$$

其中 $X_{i,j,t}$ 是零均值二阶平稳过程. 类似文献 [3], 我们用谱方法产生时空过程 $X_{i,j,t}$:

$$X_{i,j,t} = \left(\frac{2}{M} \right)^{1/2} \sum_{k=1}^M \cos(w(1, k) * i + w(2, k) * j + q(k) * t + r(k)),$$

其中 $w(i, k)$ ($i = 1, 2$) 与 $q(k)$ ($k = 1, \dots, M$) 独立同分布, 来自标准正态分布, 且独立于 $r(k)$ ($k = 1, \dots, M$), 而 $r(k)$ ($k = 1, \dots, M$) 独立同分布, 来自 $[-\pi, \pi]$ 上的均匀分布. 当 $n \rightarrow \infty$ 时, $X_{i,j,t}$ 收敛于 Gauss 遍历过程 (参见文献 [20]). 考虑以下随机误差分布的 4 种情形:

情形 1: $R_{i,j,t} \sim N(0, 1)$;

情形 2: $R_{i,j,t} \sim t_3$;

情形 3: $R_{i,j,t} \sim 0.4N(-1.4, 1) + 0.6N(0.6, 0.4^2)$;

情形 4: $R_{i,j,t} \sim 0.9N(0, 1) + 0.1N(0, 10^2)$.

情形 1 是标准正态分布, 是我们比较的基准线; 情形 2 是自由度为 3 的 t 分布; 情形 3 是左偏的; 而最后一个是混合 Gauss 分布, 10% 数据来自 $N(0, 10^2)$, 可以看作是奇异点. 对于 4 种情形, 令 $M = 1,000$, 重复模拟 100 次, 样本量大小分别为 $8 \times 8 \times 8$ 、 $10 \times 10 \times 10$ 和 $15 \times 15 \times 15$. 选取 $\mathcal{K}(\cdot)$ 为一维 E 核, $\mathcal{K}(x) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$, 其中 I 为示性函数. 为了比较局部众数方法 (local modal, LM)、最小一乘方法 (L_1) 和最小二乘方法 (L_2), 本文采用加权均方误差 (weighted average squared error, WASE) 和拟合优度 (R^2) 来衡量估计的效率. 分别定义如下:

$$\text{WASE} = \frac{1}{n^*} \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{t=1}^{n_3} [m(X_{i,j,t}) - \hat{m}(X_{i,j,t})]^2}{\text{range}(m)^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{t=1}^{n_3} [m(X_{i,j,t}) - \hat{m}(X_{i,j,t})]^2}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{t=1}^{n_3} [m(X_{i,j,t}) - \bar{m}]^2},$$

其中 $\text{range}(m)$ 和 $\bar{m}$ 分别是 $\{m(X_{i,j,t}), i = 1, \dots, n_1, j = 1, \dots, n_2, t = 1, \dots, n_3\}$ 的极差和均值. 易知, WASE 越小, R^2 越接近于 1, 估计精度越高.

表 1 和 2 是 4 种情形下 3 种方法 100 次模拟的 WASE 和 R^2 的均值. 从表 1 和 2 可以看出, 情形 1 当误差服从正态分布时, 局部众数方法的效率和最小二乘的非常接近, 都明显优于最小一乘方法. 情形 2 当误差服从重尾的 $t(3)$ 分布时, 最小一乘方法最优, 局部众数方法其次, 两者效率较接近, 最小二乘方法最差. 情形 3 是左偏的, 局部众数方法明显优于最小二乘方法和最小一乘方法. 但情形 3 左偏, 不满足对误差的两个假设条件 ($E(R_{i,t} | X_{i,t} = \mathbf{x}) = 0$ 和 (A5)), 所以可以看出局部众数方法在情形 3 的估计效果要差于其余情形. 最后情形 4 当误差服从混合 Gauss 分布时, 局部众数和最小一乘方法效率接近, 明显优于最小二乘方法. 图 1 是 3 种不同方法估计出来的曲线和真实曲线, 可以直观地看出以上类似结论. 综上可以看出, 局部众数估计方法不仅在数据有奇异点或来自重尾分布时有较好的稳健性, 而且在数据无奇异点且误差分布是正态分布时, 效率和最小二乘估计方法是渐近一致的.

3.2 实例

为了检验局部众数方法的有效性, 考虑香港新东方东地区 2000 年 1 月 1 日至 2001 年 1 月 15 日空气污染与呼吸疾病之间的关系. 主要考虑了四种污染物: 二氧化硫 (g/m^3) X_1 、可吸入颗粒物 (g/m^3) X_2 、二氧化氮 (g/m^3) X_3 、臭氧 (g/m^3) X_4 和两个环境因素: 温度 ($^{\circ}\text{C}$) X_5 和相对湿度 (%) X_6 . 其中 2000 年 2 月 19 日到 2000 年 2 月 24 日的相对湿度数据缺失, 所以这 6 天的数据被补齐. 我们感兴趣的问题是, 香港新东方东地区的化学污染物水平与呼吸系统疾病每日住院人数 (Y) 之间的关系.

由于非参数模型有维数灾难, 因此, 首先对 6 个自变量 X_1 、 X_2 、 X_3 、 X_4 、 X_5 和 X_6 作主成分分析, 根据主成分分析结果, 可以得到前 3 个主成分的累积贡献率近似为 81.5% (超过了 80%), 故取前 3 个主成分. 前 3 个主成分分别表示为

表 1 100 次模拟 WASE 的均值 (10^{-2})

	$8 \times 8 \times 8$			$10 \times 10 \times 10$			$15 \times 15 \times 15$		
	LM	L_1	L_2	LM	L_1	L_2	LM	L_1	L_2
情形 1	0.9301	1.9278	0.8221	0.6637	0.9157	0.5087	0.6289	0.8729	0.6235
情形 2	0.6665	0.4544	1.2887	0.6317	0.4169	0.9916	0.5183	0.4146	0.8177
情形 3	0.8316	3.6051	2.9927	0.8169	3.8016	3.5529	1.2321	4.7214	4.7692
情形 4	0.6622	0.9972	916.93	0.5248	0.6822	468.51	0.6720	0.6653	87.078

表 2 100 次模拟 R^2 的均值

	$8 \times 8 \times 8$			$10 \times 10 \times 10$			$15 \times 15 \times 15$		
	LM	L_1	L_2	LM	L_1	L_2	LM	L_1	L_2
情形 1	0.8914	0.7749	0.9040	0.9294	0.8931	0.9459	0.9273	0.8981	0.9279
情形 2	0.9202	0.9456	0.8457	0.9268	0.9501	0.8847	0.9418	0.9504	0.9082
情形 3	0.9048	0.5873	0.6674	0.9079	0.5630	0.5996	0.8647	0.4573	0.4763
情形 4	0.9241	0.8857	-104.1	0.9415	0.9216	-51.228	0.9226	0.9235	-9.825

图 1 (a) 情形 1: $R_{i,j,t} \sim N(0,1)$; (b) 情形 2: $R_{i,j,t} \sim t_3$; (c) 情形 3: $R_{i,j,t} \sim 0.4N(-1.4,1) + 0.6N(0.6, 0.4^2)$; (d) 情形 4: $R_{i,j,t} \sim 0.9N(0,1) + 0.1N(0,10^2)$. 样本量都为 $8 \times 8 \times 8$, 黑色实线为真实曲线 $y = 2 \sin(2\pi x)$, 黑色虚线为局部众数方法拟合的曲线, 灰色实线为最小二乘方法拟合的曲线, 灰色虚线为最小一乘方法拟合的曲线

$$Z_1 = 0.202X_1 + 0.599X_2 + 0.496X_3 + 0.249X_4 - 0.316X_5 - 0.439X_6,$$

$$Z_2 = 0.621X_1 + 0.398X_3 - 0.607X_4 + 0.126X_5 + 0.268X_6,$$

$$Z_3 = 0.407X_1 - 0.182X_3 + 0.343X_4 + 0.781X_5 - 0.254X_6.$$

此时解释变量是 3 维的, 采用乘积核

$$\mathcal{K}_{h_1}(t_1, t_2, t_3) = \frac{1}{h_1^3} K\left(\frac{t_1}{h_1}\right) K\left(\frac{t_2}{h_1}\right) K\left(\frac{t_3}{h_1}\right),$$

其中 $K(\cdot)$ 是 E 核. 依然用 WASE 和 R^2 来衡量估计的效率, 结果如表 3 所示, 可以看出在这个例子中, 局部众数方法的估计效率优于最小二乘方法.

表 3 两种方法的 WASE 和 R^2 的值

	WASE	R^2
LM	0.1208	0.7319
L_2	0.2922	0.3516

4 定理的证明

本节将给出定理的证明. 证明依赖于一些基本的结论, 以下两个引理类似于文献 [21, 22] 的结论, 我们不再给出证明.

引理 4.1 (1) 假定 (2.3) 成立, 令 $\mathcal{L}_r(\mathcal{F})$ 为所有满足条件 $\|X\|_r = (\mathbb{E}|X|^r)^{\frac{1}{r}} < \infty$ 的 $\mathcal{F}$ -可测随机变量 X 组成的集合. 令 $X \in \mathcal{L}_r(\mathcal{B}(E))$, $Y \in \mathcal{L}_s(\mathcal{B}(E'))$, 则对任意满足 $r^{-1} + s^{-1} + h^{-1} = 1$ 的 $1 \leq r, s, h < \infty$, 有

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq C\|X\|_r\|Y\|_s\{\psi_1(\text{Card}E, \text{Card}E')\varphi_1(d(E, E'))\}^{\frac{1}{h}}. \quad (4.1)$$

(2) 进一步, 若 $|X|$ 和 $|Y|$ 几乎处处有界, 则 (4.1) 的右边可以由 $C\psi_1(\text{Card}E, \text{Card}E')\varphi_1(d(E, E'))$ 代替.

引理 4.2 (1) 假定 (2.4) 成立, 令 $\mathcal{L}_r(\mathcal{F})$ 为所有满足条件 $\|X\|_r = (\mathbb{E}|X|^r)^{\frac{1}{r}} < \infty$ 的 $\mathcal{F}$ -可测随机变量 X 组成的集合. 令 $X \in \mathcal{L}_r(\mathcal{B}(G))$, $Y \in \mathcal{L}_s(\mathcal{B}(G'))$, 则对任意满足 $r^{-1} + s^{-1} + h^{-1} = 1$ 的 $1 \leq r, s, h < \infty$, 有

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq C\|X\|_r\|Y\|_s\{\psi_2(\text{Card}G, \text{Card}G')\varphi_2(d(G, G'))\}^{\frac{1}{h}}. \quad (4.2)$$

(2) 进一步, 若 $|X|$ 和 $|Y|$ 几乎处处有界, 则 (4.2) 的右边可以由 $C\psi_2(\text{Card}G, \text{Card}G')\varphi_2(d(G, G'))$ 代替.

记 $S(X_{i,t}) = m(X_{i,t}) - m(\mathbf{x}_0) - [m'(\mathbf{x}_0)]^\top (X_{i,t} - \mathbf{x}_0)$, $\mu_j = \int_{\mathbb{R}^d} \mathcal{K}(\mathbf{u}) \mathbf{u}^j d\mathbf{u}$, $j = 0, 1$.

引理 4.3 假设 (A1)–(A7) 成立, 对于 $j = 0, 1$, 有

$$\frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi''_{h_2}(R_{i,t}) \left(\frac{X_{i,t} - \mathbf{x}_0}{h_1} \right)^j = F(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mu_j + o_p(1), \quad (4.3)$$

且

$$\begin{aligned} & \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi''_{h_2}(R_{i,t}) S(X_{i,t}) \left(\frac{X_{i,t} - \mathbf{x}_0}{h_1} \right)^j \\ &= h_1^2 F(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \int_{\mathbb{R}^d} \frac{1}{2} \mathcal{K}(\mathbf{u}) \mathbf{u}^\top m''(\mathbf{x}_0) \mathbf{u} \cdot \mathbf{u}^j + o_p(h_1^2). \end{aligned} \quad (4.4)$$

证明 只证 (4.3), (4.4) 类似可证. 为了方便起见, 记

$$T_{i,t} = \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi''_{h_2}(R_{i,t}) \left(\frac{X_{i,t} - \mathbf{x}_0}{h_1} \right)^j.$$

利用文献 [23, 引理 5.1], 对于 $j = 0, 1$, 当 $n \rightarrow \infty$ 时, 有

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} T_{i,t} \right] &= \mathbb{E}[\mathbb{E}(T_{i,t} | X_{i,t} = \mathbf{x})] = \frac{1}{h_1^d} \int_{\mathbb{R}^d} \mathcal{K} \left(\frac{\mathbf{x} - \mathbf{x}_0}{h_1} \right) F(\mathbf{x}, h_2) \left(\frac{\mathbf{x} - \mathbf{x}_0}{h_1} \right)^j f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \mathcal{K}(\mathbf{u}) F(\mathbf{x}_0 + h_1 \mathbf{u}, h_2) f(\mathbf{x}_0 + h_1 \mathbf{u}) \mathbf{u}^j d\mathbf{u} \rightarrow F(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mu_j, \end{aligned} \quad (4.5)$$

而

$$\text{Var} \left[\frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} T_{i,t} \right] = \frac{1}{n^{*2}} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \text{Var}(T_{i,t}) + \frac{1}{n^{*2}} \sum_{(i,t) \neq (j,\tau)} \mathbb{E}(T_{i,t} - \mathbb{E}T_{i,t})(T_{j,\tau} - \mathbb{E}T_{j,\tau})^\top$$

$$=: A_1 + A_2,$$

其中当 $\mathbf{n} \rightarrow \infty$ 时, 由 (A3) 和 (A6), 有

$$A_1 \leq \frac{1}{n^{*2}} \sum_{\mathbf{i} \in \Lambda_{\mathbf{n}}} \sum_{t \in T_{\mathbf{n}}} \mathbb{E} T_{\mathbf{i},t} T_{\mathbf{i},t}^{\top} = \frac{1}{n^{*} h_1^d} H(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u}^j (\mathbf{u}^j)^{\top} d\mathbf{u} + o\left(\frac{1}{n^{*} h_1^d}\right) \rightarrow \mathbf{0}.$$

令 $W = \{((\mathbf{i}, t), (\mathbf{j}, \tau)) : 0 \leq \|\mathbf{i} - \mathbf{j}\| \leq c_{\mathbf{n}}, 0 \leq |t - \tau| \leq t_{\mathbf{n}}, (\mathbf{i}, t) \neq (\mathbf{j}, \tau)\}$, 其中 $c_{\mathbf{n}} = (n^{*} h_1^d)^{\frac{1}{4N}} \rightarrow \infty$, $t_{\mathbf{n}} = (n^{*} h_1^d)^{\frac{1}{4}} \rightarrow \infty$, 则

$$A_2 = \frac{1}{n^{*2}} \sum_{((\mathbf{i}, t), (\mathbf{j}, \tau)) \in W} \mathbb{E}(T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})(T_{\mathbf{j},\tau} - \mathbb{E} T_{\mathbf{j},\tau})^{\top} + \frac{1}{n^{*2}} \sum_{((\mathbf{i}, t), (\mathbf{j}, \tau)) \notin W} \mathbb{E}(T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})(T_{\mathbf{j},\tau} - \mathbb{E} T_{\mathbf{j},\tau})^{\top}.$$

因 $\text{Card}(W) \leq 2^{N+1} n^{*} c_{\mathbf{n}}^N t_{\mathbf{n}}$,

$$\mathbb{E}(T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})(T_{\mathbf{j},\tau} - \mathbb{E} T_{\mathbf{j},\tau})^{\top} \leq \mathbb{E} T_{\mathbf{i},t} T_{\mathbf{i},t}^{\top} = \frac{1}{h_1^d} H(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u}^j (\mathbf{u}^j)^{\top} d\mathbf{u} + o\left(\frac{1}{h_1^d}\right),$$

故由 $c_{\mathbf{n}}$ 和 $t_{\mathbf{n}}$ 的定义可知,

$$\begin{aligned} \frac{1}{n^{*2}} \sum_{((\mathbf{i}, t), (\mathbf{j}, \tau)) \in W} \mathbb{E}(T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})(T_{\mathbf{j},\tau} - \mathbb{E} T_{\mathbf{j},\tau})^{\top} &\leq c 2^{N+1} \frac{1}{n^{*} h_1^d} c_{\mathbf{n}}^N t_{\mathbf{n}} H(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u}^j (\mathbf{u}^j)^{\top} d\mathbf{u} \\ &\rightarrow \mathbf{0}. \end{aligned}$$

由 (4.5) 和 (A4), 有

$$\begin{aligned} h_1^d (T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t}) &= \mathcal{K}\left(\frac{X_{\mathbf{i},t} - \mathbf{x}_0}{h_1}\right) \phi_{h_2}''(R_{\mathbf{i},t}) \left(\frac{X_{\mathbf{i},t} - \mathbf{x}_0}{h_1}\right)^j - c \cdot h_1^d F(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mu_j \\ &\leq c \cdot \phi_{h_2}''(R_{\mathbf{i},t}) + o(1). \end{aligned}$$

$\phi(x)$ 取 Gauss 核, 代入得 $\phi_{h_2}''(R_{\mathbf{i},t}) = (R_{\mathbf{i},t}^2/h_2^4 - 1/h_2^2) \phi_{h_2}(R_{\mathbf{i},t})$ 有界, 所以, $h_1^d (T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})$ 有界. 令当 $j=0$ 时, $I_d=1$; 当 $j=1$ 时, I_d 为 d 维单位矩阵. 类似文献 [18, 引理 4.4] 的证明, 由引理 4.1 和 4.2, 有

$$\begin{aligned} &\frac{1}{n^{*2}} \sum_{((\mathbf{i}, t), (\mathbf{j}, \tau)) \notin W} \mathbb{E}(T_{\mathbf{i},t} - \mathbb{E} T_{\mathbf{i},t})(T_{\mathbf{j},\tau} - \mathbb{E} T_{\mathbf{j},\tau})^{\top} \\ &\leq c(n^{*} h_1^d)^{-2} \left[\sum_{\|\mathbf{i}-\mathbf{j}\| \leq c_{\mathbf{n}}} \sum_{|t-\tau| > t_{\mathbf{n}}} \varphi(\|\mathbf{i}-\mathbf{j}\|, |t-\tau|) + \sum_{\|\mathbf{i}-\mathbf{j}\| > c_{\mathbf{n}}} \varphi(\|\mathbf{i}-\mathbf{j}\|, |t-\tau|) \right] \cdot I_d \\ &\leq c \frac{2^N c_{\mathbf{n}}^N}{n^{*} h_1^{2d}} \sum_{|t| > t_{\mathbf{n}}} \varphi_2(|t|) \cdot I_d + c \frac{t_{\mathbf{n}}}{n^{*} h_1^{2d}} \sum_{\|\mathbf{i}\| > c_{\mathbf{n}}} \varphi_1(\|\mathbf{i}\|) \cdot I_d \\ &\leq c \frac{1}{n^{*\frac{3}{4}} (h_1^d)^{\frac{7}{4}}} \sum_{t > (n^{*} h_1^d)^{\frac{1}{4}}} \varphi_2(t) \cdot I_d + c \frac{1}{n^{*\frac{3}{4}} (h_1^d)^{\frac{7}{4}}} \sum_{i > (n^{*} h_1^d)^{\frac{1}{4N}}} i^{N-1} \varphi_1(i) \cdot I_d \\ &\rightarrow \mathbf{0}, \end{aligned}$$

故 $A_2 \rightarrow 0$, 由 (A1) 和 (A2), 有

$$\text{Var} \left[\frac{1}{n^{*}} \sum_{\mathbf{i} \in \Lambda_{\mathbf{n}}} \sum_{t \in T_{\mathbf{n}}} T_{\mathbf{i},t} \right] \rightarrow \mathbf{0}. \quad (4.6)$$

由 (4.5) 和 (4.6) 可知 (4.3) 成立, 故引理 4.3 得证. $\square$

记

$$X_{i,t}^* = \begin{pmatrix} 1 \\ \frac{X_{i,t} - \mathbf{x}_0}{h_1} \end{pmatrix}.$$

定理 2.1 的证明 记 $\alpha_n = (n^* h_1^d)^{-\frac{1}{2}} + h_1^2$, 为了得到定理 2.1, 只需证明对任意给定的 $\eta > 0$, 存在一个较大的 c 使得

$$P\left\{\sup_{\|\boldsymbol{\mu}=c\|} l(\boldsymbol{\theta}^* + \alpha_n \boldsymbol{\mu}) < l(\boldsymbol{\theta}^*)\right\} \geq 1 - \eta. \quad (4.7)$$

(2.1) 可以改写为

$$l(\boldsymbol{\theta}^*) = \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi_{h_2}(Y_{i,t} - \boldsymbol{\theta}^{*T} X_{i,t}^*),$$

则利用 Taylor 展式, 有

$$\begin{aligned} l(\boldsymbol{\theta}^* + \alpha_n \boldsymbol{\mu}) - l(\boldsymbol{\theta}^*) &= \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \{ \phi_{h_2}(R_{i,t} + S(X_{i,t}) - \alpha_n \boldsymbol{\mu}^\top X_{i,t}^*) \\ &\quad - \phi_{h_2}(R_{i,t} + S(X_{i,t})) \} \\ &= \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \left[-\alpha_n \boldsymbol{\mu}^\top X_{i,t}^* \phi'_{h_2}(R_{i,t} + S(X_{i,t})) \right. \\ &\quad \left. + \frac{1}{2} \alpha_n^2 (\boldsymbol{\mu}^\top X_{i,t}^*)^2 \phi''_{h_2}(R_{i,t} + S(X_{i,t})) - \frac{1}{6} \alpha_n^3 (\boldsymbol{\mu}^\top X_{i,t}^*)^3 \phi'''_{h_2}(Z_{i,t}) \right] \\ &=: B_1 + B_2 + B_3, \end{aligned}$$

其中 $Z_{i,t}$ 介于 $R_{i,t} + S(X_{i,t}) - \alpha_n \boldsymbol{\mu}^\top X_{i,t}^*$ 与 $R_{i,t} + S(X_{i,t})$ 之间.

类似引理 4.3, 由条件 (A3)、(A5) 和 Taylor 展式, 直接计算期望和方差, 有

$$\begin{aligned} E(B_1) &= -\alpha_n E\{E[\mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \boldsymbol{\mu}^\top X_{i,t}^* \phi'_{h_2}(R_{i,t} + S(X_{i,t})) \mid X_{i,t} = \mathbf{x}]\} \\ &= -\alpha_n E\{E[\mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \boldsymbol{\mu}^\top X_{i,t}^* (\phi'_{h_2}(R_{i,t}) + \phi''_{h_2}(R_{i,t}) S(X_{i,t}) + o(S(X_{i,t}))) \mid X_{i,t} = \mathbf{x}]\} \\ &= -\alpha_n \int_{\mathbb{R}^d} \frac{1}{h_1^d} \mathcal{K}\left(\frac{\mathbf{x} - \mathbf{x}_0}{h_1}\right) \boldsymbol{\mu}^\top \begin{pmatrix} 1 \\ \frac{\mathbf{x} - \mathbf{x}_0}{h_1} \end{pmatrix} F(\mathbf{x}, h_2) S(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + o(\alpha_n h_1^2) = O(c \alpha_n h_1^2), \end{aligned}$$

$$\text{Var}(B_1) = \frac{1}{n^*} \text{Var}[\mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \boldsymbol{\mu}^\top X_{i,t}^* \phi'_{h_2}(S(X_{i,t}) + R_{i,t})] = O(c^2 \alpha_n^2 (n^* h_1^d)^{-1}).$$

因此,

$$B_1 = O(c \alpha_n h_1^2) + c \alpha_n O_p((n^* h_1^d)^{-\frac{1}{2}}) = O_p(c \alpha_n^2).$$

类似可得

$$B_3 = \frac{1}{n^*} \sum_{i \in \Lambda_n} \sum_{t \in T_n} \left[-\frac{1}{6} \alpha_n^3 (\boldsymbol{\mu}^\top X_{i,t}^*)^3 \phi'''_{h_2}(Z_{i,t}) \right] = O_p(\alpha_n^3).$$

由引理 4.3, 有

$$B_2 = \alpha_n^2 F(\mathbf{x}_0, h_2) \boldsymbol{\mu}^\top \mathbf{U} \boldsymbol{\mu} (1 + o_p(1)).$$

我们知道 $\mathbf{U}$ 是正定阵, $\|\boldsymbol{\mu}\| = c$ 且 $F(\mathbf{x}_0, h_2) < 0$, 可以选择 c 充分大, 使得 B_2 至少以概率 $1 - \eta$ 控制住 B_1 和 B_3 . 因此, (4.7) 成立, 即以概率 1, 存在一个局部极大值点 $\hat{\boldsymbol{\theta}}^*$ 使得 $\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\| \leq c \alpha_n$, 则定理 2.1 成立. $\square$

定义 $H_n = \sum_{i \in \Lambda_n} \sum_{t \in T_n} X_{i,t}^* \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi'_{h_2}(R_{i,t})$, 有如下引理:

引理 4.4 假设条件 (A1)–(A7) 成立, 则有

$$\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^* = h_1^2 f(\mathbf{x}_0) \mathbf{U}^{-1} \mathbf{c}_1 + \frac{\mathbf{U}^{-1} H_n}{n^* F(\mathbf{x}_0, h_2)} (1 + o_p(1)).$$

证明 令

$$\hat{\gamma}_{i,t} = S(X_{i,t}) - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)^\top (X_{i,t} - \mathbf{x}_0) = S(X_{i,t}) - (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^\top X_{i,t}^*,$$

从而有

$$Y_{i,t} - \hat{\beta}_0 - \hat{\beta}_1^\top (X_{i,t} - \mathbf{x}_0) = R_{i,t} + \hat{\gamma}_{i,t}.$$

$\hat{\boldsymbol{\theta}}^*$ 是 (2.1) 的极大值点, 对 (2.1) 关于 $\boldsymbol{\theta}$ 求导令其等于 0, 由 Taylor 展式有

$$\begin{aligned} & \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi'_{h_2}(Y_{i,t} - \hat{\beta}_0 - \hat{\beta}_1^\top (X_{i,t} - \mathbf{x}_0)) X_{i,t}^* \\ &= \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi'_{h_2}(R_{i,t} + \hat{\gamma}_{i,t}) X_{i,t}^* \\ &= \sum_{i \in \Lambda_n} \sum_{t \in T_n} X_{i,t}^* \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \left\{ \phi'_{h_2}(R_{i,t}) + \phi''_{h_2}(R_{i,t}) \hat{\gamma}_{i,t} + \frac{1}{2} \phi'''_{h_2}(R^*) \hat{\gamma}_{i,t}^2 \right\} \\ &= 0, \end{aligned} \tag{4.8}$$

其中 R^* 介于 $R_{i,t}$ 与 $R_{i,t} + \hat{\gamma}_{i,t}$ 之间.

注意到 (4.8) 左边第二项为 $I_1 + I_2$, 其中

$$\begin{aligned} I_1 &= \sum_{i \in \Lambda_n} \sum_{t \in T_n} X_{i,t}^* \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi''_{h_2}(R_{i,t}) S(X_{i,t}), \\ I_2 &= - \sum_{i \in \Lambda_n} \sum_{t \in T_n} X_{i,t}^* \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi''_{h_2}(R_{i,t}) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^\top X_{i,t}^*. \end{aligned}$$

由引理 4.3, 有

$$I_1 = n^* h_1^2 F(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mathbf{c}_1 (1 + o_p(1)), \quad I_2 = -n^* F(\mathbf{x}_0, h_2) \mathbf{U} (1 + o_p(1)) (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*).$$

由定理 2.1 可知, $\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\| = O_p(h_1^2 + (n^* h_1^d)^{-\frac{1}{2}})$, 故

$$\begin{aligned} \sup_{(i,t): \frac{|X_{i,t} - \mathbf{x}_0|}{h_1} \leq 1} |\hat{\gamma}_{i,t}| &\leq \sup_{(i,t): \frac{|X_{i,t} - \mathbf{x}_0|}{h_1} \leq 1} [|S(X_{i,t})| + |(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^\top X_{i,t}^*|] \\ &= O_p(h_1^2 + \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|) = O_p(h_1^2 + (n^* h_1^d)^{-\frac{1}{2}}) = o_p(1) \end{aligned}$$

且

$$\sup_{(i,t): \frac{|X_{i,t} - \mathbf{x}_0|}{h_1} \leq 1} |\hat{\gamma}_{i,t}|^2 = o_p(1) O_p(\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|) = o_p(\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|). \tag{4.9}$$

类似引理 4.3 的证明, 对于 $j = 0, 1$, 有

$$\mathbb{E} \left[\frac{\mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) (X_{i,t} - \mathbf{x}_0)^j}{(h_1)^j} \right] = \int_{\mathbb{R}^d} \frac{1}{h_1^d} \frac{\mathcal{K}(\mathbf{x} - \mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^j}{(h_1)^j} f(\mathbf{x}) d\mathbf{x} = f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathcal{K}(\mathbf{u}) \mathbf{u}^j d\mathbf{u} + o(1).$$

由上式和 (4.9) 以及假设条件 (A3) 和 (A4), 有

$$\begin{aligned} & \mathbb{E} \left[\sum_{\mathbf{i} \in \Lambda_n} \sum_{t \in T_n} \frac{\mathcal{K}_{h_1}(X_{\mathbf{i},t} - \mathbf{x}_0) \hat{\gamma}_{\mathbf{i},t}^2 \phi_{h_2}'''(R^*)(X_{\mathbf{i},t} - \mathbf{x}_0)^j}{(h_1)^j} \right] \\ &= o(\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|) \sum_{\mathbf{i} \in \Lambda_n} \sum_{t \in T_n} \mathbb{E} \left[\frac{\mathcal{K}_{h_1}(X_{\mathbf{i},t} - \mathbf{x}_0)(X_{\mathbf{i},t} - \mathbf{x}_0)^j}{(h_1)^j} \right] \\ &= o(n^* \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|) = o(I_2) \end{aligned}$$

且

$$\begin{aligned} & \text{Var} \left[\sum_{\mathbf{i} \in \Lambda_n} \sum_{t \in T_n} \frac{\mathcal{K}_{h_1}(X_{\mathbf{i},t} - \mathbf{x}_0) \hat{\gamma}_{\mathbf{i},t}^2 \phi_{h_2}'''(R^*)(X_{\mathbf{i},t} - \mathbf{x}_0)^j}{(h_1)^j} \right] \\ &= o(\|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|^2) \text{Var} \sum_{\mathbf{i} \in \Lambda_n} \left[\sum_{t \in T_n} \frac{\mathcal{K}_{h_1}(X_{\mathbf{i},t} - \mathbf{x}_0)(X_{\mathbf{i},t} - \mathbf{x}_0)^j}{(h_1)^j} \right] \\ &= o(n^* \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|^2) \frac{f(\mathbf{x}_0)}{h_1^d} \int_{\mathbb{R}^d} \mathcal{K}^2(\mathbf{u}) \mathbf{u}^j (\mathbf{u}^j)^\top d\mathbf{u} \\ &= o\left(\frac{n^*}{h_1^d} \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|^2 I_d\right). \end{aligned}$$

从而, (4.8) 左端第三项为

$$\frac{1}{2} \sum_{\mathbf{i} \in \Lambda_n} \sum_{t \in T_n} X_{\mathbf{i},t}^* \mathcal{K}_{h_1}(X_{\mathbf{i},t} - \mathbf{x}_0) \hat{\gamma}_{\mathbf{i},t}^2 \phi_{h_2}'''(R^*) = o_p(I_2) + \sqrt{o_p\left(\frac{n^*}{h_1^d} \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|^2\right)} I_d = o_p(I_2).$$

再由 H_n 的定义, 有

$$\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^* = h_1^2 f(\mathbf{x}_0) \mathbf{U}^{-1} \mathbf{c}_1 + \frac{\mathbf{U}^{-1} H_n}{n^* F(\mathbf{x}_0, h_2)} (1 + o_p(1)).$$

故引理 4.4 得证. □

定理 2.2 的证明 由 (A3) 易得 $\mathbb{E}(H_n) = 0$, 类似引理 4.3 的证明, 对于 $j = 0, 1$, 有

$$\begin{aligned} \mathbb{E} \left[\frac{\mathcal{K}_{h_1}^2(X_{\mathbf{i},t} - \mathbf{x}_0) \phi_{h_2}'(R_{\mathbf{i},t})^2 (X_{\mathbf{i},t} - \mathbf{x}_0)^j}{(h_1)^j} \right] &= \int_{\mathbb{R}^d} \frac{1}{h_1^{2d}} \mathcal{K}^2\left(\frac{\mathbf{x} - \mathbf{x}_0}{h_1}\right) G(\mathbf{x}, h_2) \frac{(\mathbf{x} - \mathbf{x}_0)^j}{(h_1)^j} f(\mathbf{x}) d\mathbf{x} \\ &= h_1^{-d} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \int_{\mathbb{R}^d} \mathbf{u}^j \mathcal{K}^2(\mathbf{u}) d\mathbf{u} \{1 + o(1)\}, \end{aligned}$$

故有 $\text{Cov}(H_n) = n^* h_1^{-d} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \boldsymbol{\Sigma} (1 + o(1))$. 由引理 4.4, 易得

$$\begin{aligned} \mathbb{E}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) &= h_1^2 f(\mathbf{x}_0) \mathbf{U}^{-1} \mathbf{c}_1 (1 + o(1)), \\ \text{Var}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) &= (n^* h_1^d)^{-1} F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top. \end{aligned}$$

故定理 2.2 证毕. □

定理 2.3 的证明 由引理 4.4, 只需证

$$H_n^* = \sqrt{\frac{h_1^d}{n^*}} H_n \xrightarrow{\mathcal{L}} N(0, G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \cdot \boldsymbol{\Sigma}). \quad (4.10)$$

由 Slutsky 定理, 易得

$$(n^* h_1^d)^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^* - \mathbb{E}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)) \xrightarrow{\mathcal{L}} N(0, F(\mathbf{x}_0, h_2)^{-2} G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) \mathbf{U}^{-1} \boldsymbol{\Sigma} (\mathbf{U}^{-1})^\top).$$

下面来证 (4.10), 利用 Cramér-Wold 工具, 只需证对任意 $C \in \mathbb{R}^{d+1}$ ($C \neq \mathbf{0}$), 有

$$C^\top H_n^* \xrightarrow{\mathcal{L}} N(0, G(\mathbf{x}_0, h_2) f(\mathbf{x}_0) C^\top \boldsymbol{\Sigma} C).$$

为了简单起见, 记

$$C^\top H_n^* = \sum_{i \in \Lambda_n} \sum_{t \in T_n} \mathcal{K}_{h_1}(X_{i,t} - \mathbf{x}_0) \phi'_{h_2}(R_{i,t}) C^\top X_{i,t}^* = \sum_{i \in \Lambda_n} \sum_{t \in T_n} \Delta_{i,t}.$$

利用时空分块思想 (参见文献 [3]), 大块中所有的边等于 p_k ($k = 1, \dots, N+1$), 而小块中至少有一边等于 q . 不失一般性, 假设存在正整数 $l_1, \dots, l_{N+1}$, 使得 $n_k = l_k(p_k + q)$ 且满足 $\lim_{n \rightarrow \infty} \frac{p_k}{q} = \infty$, $k = 1, \dots, N+1$.

记

$$\begin{aligned} U(1, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k = j_k(p_k + q) + p \\ k = 1, \dots, N}}^{j_k(p_k + q) + p} \sum_{t = j_{N+1}(p_{N+1} + q) + 1}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}, \\ U(2, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k = j_k(p_k + q) + p_k \\ k = 1, \dots, N}}^{j_k(p_k + q) + p_k} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1} + 1}^{(j_{N+1} + 1)(p_{N+1} + q)} \Delta_{i,t}, \\ U(3, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N-1}}^{j_k(p_k + q) + p_k} \sum_{i_N = j_N(p_N + q) + p_N + 1}^{(j_N + 1)(p_N + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + 1}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}, \\ U(4, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N-1}}^{j_k(p_k + q) + p_k} \sum_{i_N = j_N(p_N + q) + p_N + 1}^{(j_N + 1)(p_N + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1} + 1}^{(j_{N+1} + 1)(p_{N+1} + q)} \Delta_{i,t}, \end{aligned}$$

等等. 而

$$U(2^N, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p_k + q) + p_k + 1 \\ k = 1, \dots, N}}^{(j_k + 1)(p_k + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1}}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}.$$

最终

$$U(2^{N+1}, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p_k + q) + p_k + 1 \\ k = 1, \dots, N}}^{(j_k + 1)(p_k + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1} + 1}^{(j_{N+1} + 1)(p_{N+1} + q)} \Delta_{i,t}.$$

令 $\mathcal{R} = \{0, \dots, l_1\} \times \dots \times \{0, \dots, l_{N+1}\}$, 对每个正整数 $i = 1, \dots, 2^{N+1}$, 定义

$$T(\mathbf{n}, i) = \sum_{\mathbf{j} \in \mathcal{R}} U(i, \mathbf{n}, \mathbf{j}),$$

则有

$$C^{\top} H_n^* = \sum_{i=1}^{2^{N+1}} T(n, i).$$

注意到 $T(n, 1)$ 是大块随机变量 $\Delta_{i,t}$ 的和, 而所有 $T(n, i)$ ($2 \leq i \leq 2^{d+1}$) 是小块随机变量 $\Delta_{i,t}$ 的和. 若正整数 $l_1, \dots, l_{N+1}$ 不满足 $n_k = l_k(p_k + q)$, $k = 1, \dots, N+1$, 则另添加一项 $T(n, 2^{d+1} + 1)$. 这一项中, 包含所有的既不在大块中也不在小块中的随机变量 $\Delta_{i,t}$. 添加这一项后, 证明无需作太大的改变. 类似文献 [3] 中的证明, 当 $n \rightarrow \infty$ 时, 有

$$\begin{aligned} Q_1 &= \left| E \exp[iuT(n, 1)] - \prod_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E \exp[iuU(1, n, j)] \right| \rightarrow 0, \\ Q_2 &= \left(\sum_{i=2}^{2^{N+1}} T(n, i) \right)^2 \rightarrow 0, \\ Q_3 &= \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E[U(1, n, j)]^2 \rightarrow G(x_0, h_2)f(x_0)C^{\top}\Sigma C, \\ Q_4 &= \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E[(U(1, n, j))^2 I\{|U(1, n, j)| > \varepsilon G(x_0, h_2)f(x_0)C^{\top}\Sigma C\}}] \rightarrow 0. \end{aligned}$$

故 (4.10) 成立, 即定理 2.3 得证. □

致谢 感谢审稿人对本文提出的宝贵意见.

参考文献

- 1 Hallin M, Lu Z D, Tran L T. Local linear spatial regression. *Ann Statist*, 2004, 32: 2469–2500
- 2 Biau G, Cadre B. Nonparametric spatial prediction. *Stat Inference Stoch Process*, 2004, 7: 327–349
- 3 Wang H X, Wang J D. Estimation of the trend function for spatio-temporal models. *J Nonparametr Stat*, 2009, 21: 567–588
- 4 Wang L H. The nonparametric estimation of long memory spatio-temporal random field models. *Sci China Math*, 2015, 58: 1115–1128
- 5 Lu Z D, Tang Q G, Cheng L S. Estimating spatial quantile regression with functional coefficients: A robust semiparametric framework. *Bernoulli*, 2014, 20: 164–189
- 6 Wang H X, Lin J G, Wang J D. Local linear estimation for spatiotemporal models based on least absolute deviation. *Comm Statist Theory Methods*, 2015, 44: 1508–1522
- 7 Yao W X, Lindsay B G, Li R Z. Local modal regression. *J Nonparametr Stat*, 2012, 24: 647–663
- 8 Liu J G, Zhang R, Zhao W, et al. A robust and efficient estimation method for single index models. *J Multivariate Anal*, 2013, 122: 226–238
- 9 Zhao W, Zhang R, Liu J G, et al. Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Ann Inst Statist Math*, 2014, 66: 165–191
- 10 Zhu H M, Lv Z K, Yu K M, et al. Robust variable selection in partially varying coefficient single-index model. *J Korean Statist Soc*, 2015, 44: 45–57
- 11 Lv Z K, Zhu H M, Yu K M. Robust variable selection for nonlinear models with diverging number of parameters. *Statist Probab Lett*, 2014, 91: 90–97
- 12 Zhou H, Huang X. Nonparametric modal regression in the presence of measurement error. *Electron J Stat*, 2016, 10: 3579–3620
- 13 Wang X, Chen H, Shen D, et al. Regularized modal regression with applications in cognitive impairment prediction. *Adv Neural Inf Process Syst*, 2017, 30: 1448–1458

- 14 Chen Y C, Genovese C R, Tibshirani R J, et al. Nonparametric modal regression. *Ann Statist*, 2016, 44: 489–514
- 15 Chen Y C. Modal regression using kernel density estimation: A review. *Comput Statist*, 2018, 10: e1431
- 16 Cressie N A C, Wikle C K. *Statistics for Spatio-Temporal Data*. New York: John Wiley & Sons, 2011
- 17 Li J, Ray S, Lindsay B G. A nonparametric statistical approach to clustering via mode identification. *J Mach Learn Res*, 2007, 8: 1687–1723
- 18 Wang H X, Wang J D, Huang B. Prediction for spatio-temporal models with autoregression in errors. *J Nonparametr Stat*, 2012, 24: 217–244
- 19 Cleveland W S. Robust locally weighted regression and smoothing scatterplots. *J Amer Statist Assoc*, 1979, 74: 829–836
- 20 Cressie N A C. *Statistics for Spatial Data*. New York: John Wiley & Sons, 1993
- 21 Ibragimov I A, Linnik Y V. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Nordhoff, 1971
- 22 Deo C M. A note on empirical process of strong mixing sequences. *Ann Statist*, 1973, 1: 870–875
- 23 Fan J, Jiang J. Variable bandwidth and one-step local M-estimator. *Sci China Ser A*, 2000, 43: 65–81

Local modal regression for the spatio-temporal model

Hongxia Wang, Jinguan Lin & Xingfang Huang

Abstract When the data contain outliers or come from population with heavy-tailed distributions, which appear very often in spatio-temporal data, the estimation methods based on the least square method will not perform well. More robust estimation methods are required. We propose the local linear estimation for the spatio-temporal model based on the local modal method. Asymptotic theory properties and data analysis results show that the proposed estimator is more efficient than the ordinary least square-based estimation in the case of outliers or heavy-tailed error distributions, and as asymptotically efficient as the least square estimator when there are no outliers and the error is a normal distribution. The modal expectation-maximization algorithm is adopted and the asymptotic distributions of estimators are driven when the data are mixing correlation.

Keywords spatio-temporal model, modal expectation-maximization, mixing correlation, local linear regression

MSC(2020) 62G05, 62G08

doi: 10.1360/N012018-00191